

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
15 May 2003 (15.05.2003)

PCT

(10) International Publication Number  
**WO 03/040310 A2**

- (51) International Patent Classification<sup>7</sup>: **C12N**
- (21) International Application Number: **PCT/US02/33907**
- (22) International Filing Date: 22 October 2002 (22.10.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
10/000,840 23 October 2001 (23.10.2001) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:  
US 10/000,840 (CIP)  
Filed on 23 October 2001 (23.10.2001)
- (71) Applicant (for all designated States except US):  
**WARATAH PHARMACEUTICALS, INC.** [CA/CA];  
415 Yonge Street, Suite 1103, Toronto, Ontario M5B 2E7 (CA).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **MAGIL, Sheila, G.** [US/US]; 10 Upland Road, Arlington, MA 02474 (US).

**JONES, Susan, D.** [US/US]; 7 Moon Hill Road, Lexington, MA 02421 (US). **PACE, Gary, W.** [AU/US]; 6 Robinson Park Circle, Winchester, MA 01890 (US). **BRAND, Stephen, J.** [AU/US]; 161 Bedford Road, Lincoln, MA 01773 (US).

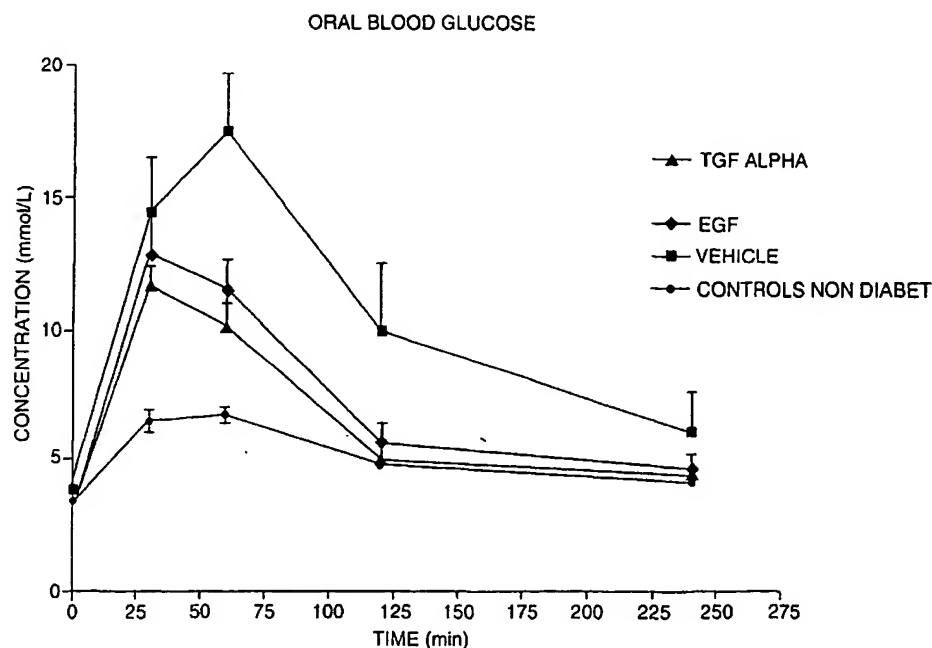
(74) Agent: **GUTERMAN, Sonia, K.**; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: NOVEL EPIDERMAL GROWTH FACTOR PROTEIN AND GENE, AND METHODS OF USE THEREFOR



(57) Abstract: Forms of epidermal growth factor that are resistant to proteolysis, and gene sequences encoding these forms and having codons optimized for usage by an industrial production organism, are provided.

WO 03/040310 A2



**Published:**

— without international search report and to be republished  
upon receipt of that report

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

**NOVEL EPIDERMAL GROWTH FACTOR PROTEIN AND GENE, AND  
METHODS OF USE THEREFOR  
FIELD OF THE INVENTION**

The present invention relates to epidermal growth factor (EGF) protein sequences having increased resistance to proteolysis and equivalent potency to normal human EGF, and to gene sequences having optimal codon usage in industrial production organisms.

**BACKGROUND**

Full length, wild-type human epidermal growth factor (EGF; see SEQ ID NO: 1) is a 53 amino acid protein with a molecular weight of 6217 daltons and a variety of biological functions (Karnes, W., *Epidermal growth factor and transforming growth factor alpha*, 1994, Raven Press, New York). Modifications of the amino acid sequence at the C-terminus have been reporting both from construction of altered forms by recombinant DNA engineering genetic studies, and as observed with EGF isolated from nature. EGF is susceptible to both endo and exo proteases, and proteolytic attack occurs in the stomach on EGF produced in the salivary gland and swallowed, has been observed as well as with EGF in the blood stream (Araki et al., *Chem. Pharm. Bull.* 37(2), 404-406, 1989; Playford et al., *Gastronenterology*, 108, 92-101, 1996).

EGF that is equal to or less than 46 amino acids in length has lost substantial biological activities, in comparison to EGF species of chain lengths equal to or greater than 47 amino acids. There are contradictory data on the exact effect on chain length on functional biological activities such as affinity for receptors and *in vivo* rates of EGF clearance. EGF of lengths 47, 48 and 51 amino acids (indicated "EGF47", "EGF48", etc.) can inhibit stomach acid secretion (U.S. patent number 3,917,824), for example, human EGF47 inhibits acid secretion with the same potency as EGF53, however this protein retains only about one tenth of the potency to stimulate fibroblast growth (mitogenic activity; Hollenberg et al., *Molecular Pharmacology* 17, 314-320, 1980; Gregory et al., *Regulatory Peptides* 22, 217-226, 1988). As these references show, the fact that a composition demonstrates high biological activity with respect to one biological activity does not imply that all biological activities are present in amounts equipotent to the full length composition.

EGF52 is equipotent to EGF53, for both inhibition of acid secretion and stimulation of cell proliferation. The mitogenic activity of mouse EGF is largely lost if the chain length is less than 48 amino acids (Burgess et al., *Biochemistry* 27, 4977-4985, 1988). Further, hEGF51 and EGF53 display similar pharmacokinetics (Kuo et al., *Drug Metabolism and Disposition* 20, 23-30, 1992). EGF51 has similar activities as EGF53 (Calnan et al., *Gut* 47,

622-627, 2000), except for the retention of immunosuppressive activity (Koch et al., *J. Molecular Biochemistry* 25, 45-59, 1984). EGF48 is reported to be stable to proteolysis and to retain biological activities (U.S. patent number 5,434,135; Kuo et al. op.cit; Sizemore et al. *Peptides* 17, 1229-1236, 1996). However, EGF48 has significantly lower activity than EGF 53 (Goodlad et al., *Clinical Science*, 91, 503-507, 1996).

Correct formation of disulfide bonds in EGF and its biological activity are not affected by shortening of the N-terminus by up to 5 amino acids (Shin, S. et al., *Peptides* 16, 205-210, 1995; DiAugustine et al., *Analytical Biochemistry* 165, 420-429, 1987). Oxidation of the methionine residue at position 21 does not affect the biological activity of recombinant h-EGF produced by yeast (George-Nascimento et al. *Biochemistry*, 27, 797-802).

Recombinant EGF is degraded by microbial proteases during production. Recombinant hEGF53 produced in *Saccharomyces cerevisiae* is degraded to a sequence of 52 and then to 51 amino acids in length, as a result of protease activity during fermentation (George-Nascimento, *Biochemistry*, 27, 797-802, 1988). EGF produced by *Pichia pastoris* is degraded to a form having 48 amino acids that is stable, and which is described as retaining high biological activity. (U.S. patent number 5,102,789). Similarly, mouse EGF produced and secreted by *Pichia pastoris* is partly degraded during fermentation to 51 amino acids in length (Clare et al., *Gene* 105, 205-212, 1991).

Comparison of these data raises questions regarding susceptibilities to proteolysis of various forms of EGF, and correlations between the lengths of these different forms and the extent of several different biological activities. Because biological activities are mediated by a family of receptors that can be differently expressed, i.e., are tissue specific, any relevant biological activity of a modified EGF must be tested for that activity, to ascertain its level of function, since a biological activity level is not necessarily predictable from data obtained using an assay of another biological activity.

EGF and EGF receptor ligands such as TGF $\alpha$  have been shown to comprise a treatment for diabetes (Nardi et al., U.S. patent numbers 5,885,956, issued March 23, 1999; and Nardi et al., 6,288,301, issued Sept. 11, 2001), a disease that has achieved epidemic proportions in the United States and elsewhere.

There is a need for EGF proteins that are stable to proteolysis, that are produced as a single molecular species in high yields in a safe and convenient production organism with reproducible composition and purity required for approval as a drug, and that retain substantial biological activity for such a therapeutic purpose.

## SUMMARY OF THE INVENTION

The present invention features a composition comprising an amino acid sequence of length X, X being an integer that is at least 48 and not more than 53, such sequence (i) being substantially homologous to a portion of SEQ ID NO: 1 from position 1 to position X-1 of SEQ ID NO: 1, and (ii) having at position X an amino acid residue different from that found in SEQ ID NO: 1. In accordance with a related embodiment, the amino acid residue at position X is a neutral amino acid. The amino acid residue at position X is, for example, asparagine. Further, X is position 51. The composition has increased resistance to proteolysis in comparison with that of SEQ ID NO: 1. The biological activity is at least 75% of that of SEQ ID NO: 1, for example, the biological activity is at least 90% of that of SEQ ID NO: 1. The biological activity is selected from the group consisting of: mitogenesis, cytoprotection, inhibition of acid secretion, growth of a tissue precursor cell, differentiation of a tissue precursor cell, and growth and differentiation of a tissue precursor cell. Mitogenesis is determined by the effect of the amino acid sequence on rate of mitosis of epithelial cells. Acid secretion is determined in a gastric fistulated animal. Differentiation is determined by islet neogenesis or mucosal cell formation.

In accordance with related embodiments, X is a hydrophobic amino acid, X is a neutral amino acid, or X is a charged amino acid. In a related embodiment when X is a neutral amino acid, X can be selected from glutamine, alanine, and serine.

In another aspect, the invention provides a composition comprising an amino acid sequence substantially homologous to SEQ ID NO: 1, wherein the amino acid residue at position 51 of SEQ ID NO: 1 is an amino acid other than glutamic acid. The biological activity of the composition is at least 50% of that of SEQ ID NO: 1. For example, the biological activity of the composition is at least 75% of that of SEQ ID NO: 1, for example, the biological activity of the composition is at least 90% of that of SEQ ID NO: 1. The composition has a biological activity substantially equivalent to that of SEQ ID NO: 1. In accordance with a related embodiment, the amino acid residues at positions 1-50 are at least 75% identical to that of SEQ ID NO: 1.

In another aspect, the invention provides a composition comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 2, 3 and 4. For example, the invention in one embodiment provides an amino acid sequence composition as shown in SEQ ID NO: 2. Further, the invention in another embodiment provides a polypeptide of 51 amino acids in length, wherein residues 1-50 are substantially homologous to the amino acid

sequence as shown in SEQ ID NO: 1, and residue 51 is an asparagine residue. The polypeptide has increased resistance to proteolysis in comparison with that of SEQ ID NO: 1.

In a related embodiment, the invention provides a polypeptide wherein at least one of residues 1-50 is a conservative substitution of an amino acid in the sequence as shown in SEQ ID NO: 1. The polypeptide in a related embodiment further comprises a deletion of at least one of residues selected from positions 1-5 as shown in SEQ ID NO: 1. The polypeptide has at least 50% of a biological activity of human EGF as shown in SEQ ID NO: 1.

In another aspect, the invention provides a composition comprising a human epidermal growth factor (EGF) having an amino acid sequence substantially homologous to that of at least positions 1-47 as shown in SEQ ID NO: 1, and having at least one amino acid replacement at positions 48-53 of the EGF carboxy terminus, the amino acid sequence being more stable to proteolysis than that of SEQ ID NO: 1. In a related embodiment, the amino acid sequence for residues at positions 1-50 is substantially as shown in SEQ ID NO: 1, and the residue at position 51 is an amino acid other than glutamic acid. The residue in a related embodiment at position 51 is selected from the group consisting of asparagine, glutamine, alanine, and serine, for example, the residue at position 51 is asparagine. At least 75% of the amino acids at positions 1-50 are as shown in SEQ ID NO: 1. The biological activity of the composition is at least 50% of that shown in SEQ ID NO: 1. The biological activity is selected from the group consisting of mitogenesis, cytoprotection, inhibition of acid secretion, growth of a tissue precursor cell, differentiation of a precursor cell, and growth and differentiation of a precursor cell. Mitogenesis is determined by rate of mitosis of epithelial cells. Acid secretion is determined in a gastric fistulated animal. Differentiation is determined by islet neogenesis or mucosal cell formation.

In another embodiment, the invention provides a pharmaceutical composition comprising an effective dose of any of the compositions herein, in a pharmaceutically acceptable carrier. The pharmaceutical composition in a related embodiment, further comprises an additional therapeutic agent. For example, the additional therapeutic agent is a growth factor receptor ligand. For example, the ligand is a growth factor. For example, the ligand is a gastrin/cholecystokinin receptor ligand.

In a related embodiment, the invention provides a nucleotide sequence encoding an EGF composition herein, the nucleotide sequence having codons adjusted for optimum usage in an industrially acceptable production organism. The industrially acceptable production organism is a yeast. For example, the yeast is *Pichia pastoris*.

In a related aspect, the invention provides a polynucleotide having a nucleotide sequence encoding a polypeptide of 51 residues in length and having a biological activity of human EGF, the sequence containing codons that are optimized for expression in a species of *Pichia*, and having an amino acid at the carboxyl terminus capable of conferring resistance to proteolysis. In related embodiments, the invention provides a recombinant strain of *Pichia* carrying a nucleotide sequence as shown in SEQ ID NO: 6; a recombinant strain of *Pichia* capable of producing the amino acid sequence as shown in SEQ ID NO: 2; and a nucleotide sequence encoding an amino acid sequence as shown in SEQ ID NO: 2.

In another aspect, the invention provides a method of obtaining a composition comprising an amino acid sequence of length X, X being an integer that is at least 48 and not more than 53, such sequence (i) being substantially homologous to the portion of SEQ ID NO: 1 from position 1 to position X-1 of SEQ ID NO: 1, and (ii) having at position X an amino acid residue different from that found in SEQ ID NO: 1, the method comprising: designing a gene encoding the composition, having codons selected for optimum usage in an industrially acceptable organism; and producing the composition in the organism by fermentation of the organism.

A related aspect is a method of manufacture of a therapeutic composition for use to treat a subject in need of regeneration of a tissue, a therapeutic composition comprising an amino acid sequence of length X, X being an integer that is at least 48 and not more than 53, such sequence (i) being substantially homologous to the portion of SEQ ID NO: 1 from position 1 to position X-1 of SEQ ID NO: 1, and (ii) having at position X an amino acid residue different from that found in SEQ ID NO: 1; and administering to the subject the composition in an amount sufficient to effect regeneration of precursor cells, as a treatment of the subject for regeneration of the tissue. Accordingly, administering the composition is further administering an additional therapeutic agent. For example, the additional therapeutic agent is a growth factor receptor ligand. For example, the ligand is a growth factor. For example, the ligand is a gastrin/cholecystokinin receptor ligand. In related embodiments, X is the integer 51. The amino acid residue at X is a neutral amino acid. For example, the amino acid residue is asparagine. The composition is more resistant to proteolysis than the composition having a sequence as shown in SEQ ID NO: 1. In related embodiments, the subject is in need of islet cell regeneration, for example, subject has diabetes; or the subject is in need of mucosal cell regeneration.

In another aspect, the invention provides a method of obtaining a modified amino acid sequence of an EGF that is more resistant to proteolysis than a nature identical EGF and that

substantially retains a biological activity, the method comprising: identifying at least one residue of a sequence of the nature identical EGF that is subject to proteolysis by an industrially useful organism; designing the modified amino acid sequence in which the residue identified as subject to proteolysis is deleted or substituted with a different amino acid; and providing the modified amino acid sequence and testing proteolysis in comparison to the nature identical EGF, to obtain a modified EGF that is more resistant to proteolysis than the nature identical EGF. Accordingly, the nature identical EGF is from a human. Designing the modified amino acid sequence is, in a related embodiment, substituting at least one amino acid for at least one carboxy terminal residue, to obtain a new carboxy terminal amino acid sequence, such that the modified EGF is more resistant to proteolysis. In a related embodiment, identifying the residue that is subject to proteolysis is determining a site of proteolysis during production of the nature identical EGF in a culture of a recombinant cell of the industrially useful organism. In a related method, following designing the modified amino acid sequence and prior to providing the modified EGF, the method further includes designing a nucleotide sequence encoding the modified amino acid sequence, the nucleotide sequence having codons selected for optimal usage in the industrially useful organism. In a related method, following designing the modified nucleotide sequence and prior to providing the modified EGF, the method further includes incorporating the modified nucleotide sequence into a vector, and transforming the vector into a cell of the industrially useful organism. In one embodiment, providing the modified EGF is providing a protein having an amino acid sequence as shown in SEQ ID NO:2. The industrially useful organism is a fungus, for example, is a yeast such as *Pichia*, for example, *P. pastoris*. Other industrially useful fungal organisms include species of *Neurospora*, *Aspergillus*, *Saccharomyces*, *Torula*, and *Schizosaccharomyces*. In another aspect, the invention provides a recombinant strain of *Pichia* obtained according to the previous method. In a related embodiment, the useful organism is a bacterium for example, the bacterium is a strain of *Streptomyces*, for example, *S. lividans*, *S. coelicolor*, *S. rimosus*, or other useful strains of actinomycetes such as *Actinomadura* and *Nocardia*. Other industrially useful non-actinomycete bacterial species include *Bacillus*, *Escherichia*, and *Xanthobacter*.

The invention also provides a kit comprising at least one unit dosage of the composition of a composition herein in a pharmaceutically acceptable carrier. The kit according to a related embodiment further comprises an additional therapeutic agent. For example, the additional therapeutic agent is a growth receptor ligand; for example, the growth receptor ligand is a gastrin/cholecystokinin receptor ligand, for example, the ligand is gastrin.



The unit dosage is sufficient for treatment of a subject in need of regeneration of a tissue. The tissue is a pancreatic islet or a gastric mucosa.

In another aspect, the invention provides a transgenic animal carrying the nucleotide sequence encoding a composition herein having codons adjusted for optimum usage in an industrially acceptable organism, the industrially acceptable organism being the animal. In a related embodiment, the nucleotide sequence contains additional regulatory sequence information for expression in a specific tissue. The animal can further contain an insertion mutation in the gene encoding nature identical endogenous EGF, such that production of the endogenous EGF of the animal is effectively knocked out.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 panels A, B, C, and D show, respectively, amino acid sequences and nucleotide sequences encoding proteins and genes for EGF51N (SEQ ID NO: 2), EGF51A (SEQ ID NO: 3), EGF51Q (SEQ ID NO: 4), and EGF51S (SEQ ID NO: 5), which are modified forms of EGF, each having a chain length of 51 amino acids and a C-terminus residue which is asparagine, alanine, glutamine, or serine, respectively. The encoding genes are SEQ ID NOs: 6, 7, 8, and 9, respectively.

Figure 2 shows production of EGF51N by *Pichia pastoris*. The amounts produced of each of three isoforms of amino acid residue lengths 49, 51, and 50, respectively, indicated as A, B, and C, and are shown as a function of time during the production, as is total amount of hEFG. Little to none of the A and C forms (49 or 50 amino acid residues in length) were observed, while EGF of 51 residues in length was continuously and stably produced as a function of growth of cells.

Figure 3 shows results of treatment of streptozotocin-induced diabetic rats with 40 µg/kg/day of each of human gastrin and EGF51N, delivered intraperitoneally by continuous infusion for 14 days. After treatment, rats were evaluated for ability to restore plasma glucose to resting levels, shown on the ordinate, as a function of time in minutes shown on the abscissa, after an oral glucose challenge. Rapid restoration to resting levels of plasma glucose indicates biological activity of EGF51N.

#### DESCRIPTION OF SPECIFIC EMBODIMENTS

Human EGF was discovered in human urine and was named urogastrone due to its ability to inhibit acid secretion in the stomach (Gregory, H. et al. *Hoppe Seylers.Z.Physiol Chem.* 356, 1765-1774, 1975). In addition, EGF has mitogenic activity, i.e., it stimulates the growth of various cells and tissues (Karnes *supra*; Carpenter, G. et al. *J.Cell Physiol* 88, 227-237, 1976; Gasslander, T. et al. *Eur.Surg.Res.* 29, 142-149, 1997). EGF is also found to have

a cytoprotective effect, stimulating migration of a cell toward a wound *in vivo*, or toward a gap introduced in a monolayer of cells in culture, to promote wound healing. These biological activities are specific to a family of structurally related growth factors, including EGF, TGF- $\alpha$ , amphiregulin and heparin binding EGF-like growth factor (Karnes, *supra*). The members of this family of growth factors have identical amino acids at 11 residues of the amino acid sequence, six of which are cysteine residues that form disulfide bonds.

Full length, natural (normal or wild-type) human epidermal growth factor (EGF) is a 53 amino acid protein with a molecular weight of 6217 daltons, the protein having a variety of biological functions *in vivo* and *in vitro* (Karnes, W., *Epidermal growth factor and transforming growth factor alpha*, 1994, Raven Press, New York). The term "natural EGF" as used herein shall mean full length, normal human EGF, as shown in SEQ ID NO: 1. The term "epidermal growth factor" or "EGF", as used throughout the specification and in the claims, refers to a polypeptide product or pharmaceutically acceptable salt thereof, which exhibits biological activities that are similar to natural human epidermal growth factor (hEGF; SEQ ID No: 1), as measured in one or more bioassays.

EGF receptor ligands include a family of proteins, including EGF and TGF $\alpha$ , capable of binding to a variety of EGF receptors on cells on various cell types in different tissues, and transmitting a signal to those cells, causing changes in growth and development of the particular cell type.

A number of forms of "modified EGF", varying from natural EGF in chain length and amino acid sequence, have been engineered and characterized as described herein. These modifications have been shown to affect both a biological activity and the rate of clearance of EGF. Further, the term includes peptides having the same or a similar amino acid sequence as hEGF, for example, with conservative amino acid substitutions at various residues. One or more of the last 5 amino acids from the C-terminus can be substituted with one or more other amino acids, or can be deleted.

Recombinant EGF having a methionine at position 21 replaced by a leucine residue has been described (U.S. patent number 4,760,023). Recombinant hEGF was converted during storage from an aspartyl residue at position 11, to an isoaspartyl form that showed greatly reduced biological activity (George-Nascimento et al., *Biochemistry*, 29, 9584-9591, 1990). A series of nucleic acid molecules have been described that encode a family of protein that have significant similarity to EGF and TGF- $\alpha$  (WO 00/29438). EGF muteins (mutated EGF) having histidine at residue 16 replaced with a neutral or acidic amino acid have been described (WO 93/03757), such forms retaining activity at low values of pH.

Chemical analogues and fragments of EGF and TGF- $\alpha$  retain ability to bind various members of the EGF receptor family (U.S. patent number 4,686,283). Further, full length and other forms of EGF are susceptible in vivo and in vitro both to oxidation and to proteolysis.

Embodiments of the present invention are based on the discovery that certain modifications to a C-terminus amino acid sequence of EGF, the C-terminus ranging from amino acid residue at position 48 to position 53, can result in forms of EGF that are resistant to endo- and exo-protease activity, and that retain full biological activities. These include EGF forms in which amino acids are deleted or replaced, for example, the basic amino acids at positions 48 (lysine in natural EGF) and 53 (arginine), the aromatic amino acids at positions 49 (tryptophan) and 50 (tryptophan), and the aliphatic amino acid at position 52 (leucine). The embodiments herein include pharmaceutically acceptable salts of the modified forms of EGF herein.

A "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antimicrobials such as antibacterial and antifungal agents, isotonic and absorption delaying agents and the like that are physiologically compatible. Preferably, the carrier is suitable for intravenous, intramuscular, oral, intraperitoneal, transdermal, or subcutaneous administration. See, "Controlled Release of Drugs: Polymers and Aggregate Systems", M. Rosoff, Ed., John Wiley, Inc., NY (1989).

As used herein, the term "pharmaceutically acceptable salt" refers to a salt that retains the desired biological activity of the parent compound and does not impart any undesired toxicological effects. Examples of such salts are (a) acid addition salts formed with inorganic acids, for example hydrochloric acid, hydrobromic acid, sulfuric acid, phosphoric acid, nitric acid and the like; and salts formed with organic acids such as, for example, acetic acid, oxalic acid, tartaric acid, succinic acid, maleic acid, fumaric acid, gluconic acid, citric acid, malic acid, ascorbic acid, benzoic acid, tannic acid, pamoic acid, alginic acid, polyglutamic acid, naphthalenesulfonic acids, naphthalenedisulfonic acids, polygalacturonic acid; (b) salts with polyvalent metal cations such as zinc, calcium, bismuth, barium, magnesium, aluminum, copper, cobalt, nickel, cadmium, and the like; or (c) salts formed with an organic cation formed from N,N'-dibenzylethylenediamine or ethylenediamine; or (d) combinations of (a) and (b) or (c), e.g., a zinc tannate salt; and the like.

A physician or one having ordinary skill in the art can readily determine and prescribe an "effective dose" of the pharmaceutical composition required. For example, the physician could start administering a dose of the compound of the invention in the pharmaceutical composition at a level lower than that required in order to achieve the desired therapeutic

effect, e.g., remediation of diabetes type I or type II or streptozotocin induced diabetes, and increase the dosage with time, until the desired effect is achieved, i.e., remediation of diabetes.

Dosage regimens can be adjusted to provide the optimum desired response, e.g., a therapeutic response, specifically herein remediation of a form of diabetes. A single dosage such as a single bolus can be administered, several divided doses can be administered over time, or the dose can be proportionally reduced or increased as indicated by the exigencies of the disease situation. A physician or other practitioner having ordinary skill in the pharmacological arts can readily determine and prescribe the effective dose of the required pharmaceutical composition. For example, the practitioner could start administering doses at a level lower than that required to achieve the desired therapeutic effect, and increase the dosage with time to obtain the desired effect, specifically, mitigation of symptoms of diabetes type I, type II, or streptozotocin induced diabetes. In general, a suitable effective dose of a modified EGF composition will be the lowest dose producing mitigation of symptoms such as mitigation of failure to respond to a glucose challenge by production of insulin and reduction in blood sugar concentration.

In another embodiment, a pharmaceutical composition herein includes also an additional therapeutic agent. Thus according to a method herein, a pharmaceutical composition which comprises a modified EGF can be administered as part of a combination therapy, in combination with an additional agent or agents, for example, in combination with a cholecystokinin receptor ligand.

A therapeutically effective dosage reduces symptomology by at least about 20%, at least about 40%, at least about 60%, or at least about 80%, compared to untreated subjects that have not received the composition.

The amino acids that occur in the various amino acid sequences referred to in the specification shall have their usual, three- and one-letter abbreviations, routinely used in the art. "Hydrophobic" amino acids include the aromatic amino acids, tyrosine, tryptophan, and phenylalanine, and the aliphatic amino acids isoleucine, leucine, and valine. "Charged" amino acids include the acidic amino acids glutamic acid and aspartic acid, and the basic amino acids lysine and arginine. Other amino acids which are not charged are "neutral" amino acids.

A "conservative" amino acid substitution mutation shall mean that an amino acid found at a particular position in an EGF or a related molecule is replaced by one that is chemically similar. Examples of conservative amino acid substitution are: a charged amino

acid replaced by a different amino of the same charge, such as Asp replaced by Glu, or an aromatic hydrophobic amino acid, e.g., Trp, replaced by a different aromatic hydrophobic amino acid, e.g., Phe (see U.S. patent number 6,207,154, issued March 27, 2001). A modified EGF, e.g., EGF51N, having in addition one or more conservative substitutions of amino acids, is considered to be encompassed within embodiments that are equivalent in the meaning of the various modified EGF composition, as described and claimed herein.

An amino acid sequence composition is "substantially homologous" to another if a major percentage of the particular type of amino acid residues that make up the sequence of the first are identical to the amino acids in the sequence of the other, each lined up linearly so that comparisons can be made on a position-by-position basis. For example, at least 50%, at least 75%, at least 85%, at least 90% or at least 95% of the amino acids in the sequence are identical. For a protein of 51 residues in length compared to another of that length, at least 50% means at least 26 amino acids of the sequences must be identical; 75% means at least 39 amino acids of the sequences must be identical; 85% means at least 43 amino acids of the sequences must be identical; 90% means at least 46 must be identical; and 95% means at least 48 amino acids of the sequence must be identical to that of the first sequence in the comparison.

**Table 1. Three letter and one letter amino acid abbreviations**

Amino Acid	Abbreviation	
L-Alanine	Ala	A
L-Arginine	Arg	R
L-Asparagine	Asn	N
L-Aspartic acid	Asp	D
L-Cysteine	Cys	C
L-Glutamine	Gln	Q
L-Glutamic Acid	Glu	E
L-Glycine	Gly	G
L-Histidine	His	H
L-Isoleucine	Ile	I
L-Leucine	Leu	L
L-Lysine	Lys	K
L-Methionine	Met	M
L-Phenylalanine	Phe	F
L-Proline	Pro	P
L-Serine	Ser	S
L-Threonine	Thr	T
L-Tryptophan	Trp	W
L-Tyrosine	Tyr	Y
L-Valine	Val	V

One activity of EGF recently shown by Nardi et al. (U.S. patent numbers 5,885,956, and 6,288,301) is that administration to a subject of an EGF receptor ligand in combination with a gastrin/cholecystokinin (CCK) receptor ligand enables a pancreatic islet precursor cell in the subject to differentiate, and to mature to an insulin-secreting cell. An EGF receptor ligand can thus play a role in causing differentiation of an undifferentiated progenitor cell, alone or in combination with another agent such as a CCK receptor ligand. A "progenitor" cell has the capability to divide for several generations, and to differentiate into a variety of different cell types.

EGF and other EGF receptor ligands have a mitogenic activity, and are capable of stimulating proliferation of cell number, particularly of an epithelial cell line in culture or in vivo. Yet another activity of an EGF receptor ligand is suppression of acid secretion, as demonstrated in an experimental animal, such as a fistulated animal, for example, a fistulated rat. EGF can further stimulate either or both of growth and differentiation of a tissue precursor cell.

The modified EGF compositions herein can be further modified to incorporate a chemical analog at one or more amino acid positions, for the purpose of inhibiting additional proteolysis that might shorten the pharmacological effectiveness of the modified EGF in vivo. While the modified EGF compositions as shown herein are demonstrably more resistant to proteolysis than is nature identical hEGF, it is anticipated that further desirable chemical change is envisioned by embodiments of the invention described herein. Such further modifications include, for example, the presence of at least one D-amino acid substituted for a natural L-amino acid at a particular position, or the substitution of at least one alanine or another amino acid residue with a compound such as norvaline, acetylcysteine, or methylphenylglycine. An amino acid modification can also be N-methylation of a peptide backbone nitrogen.

The term "proteolysis" as used herein shall mean the process by which a protease or peptidase hydrolyzes a peptide bond, in the group of enzymes known as hydrolases. Proteolysis as used herein and in the claims includes both exopeptidase processive activities that sequentially release single amino acids from a terminus of the substrate protein, and endoprotease cleavage activities that produce two or more polypeptide, oligopeptide, or amino acid fragments following digestion of the substrate.

An "industrially useful organism" shall mean a strain of a microorganism or a cell line that is generally regarded as safe, in which a therapeutic agent can be produced for

administration to a human or animal, such that the organism does not contribute additional molecular entities that would provoke negative side effects or sequelae. A variety of bacteria have been used for industrial production, such as *Streptomyces* (see U.S. patent number 4,745,056, issued May 17, 1988). For proteins that are sensitive to proteolytic degradation, species of fungi, for example, yeasts such as *Saccharomyces* species, for example, *S. cerevisiae*, and species of *Pichia*, for example, *P. pastoris*, are particularly useful for robust growth and production of a therapeutic protein, in the absence of significant contribution of other undesired molecular entities.

Further, the modified EGF forms as described herein, for example, EGF51N, can be manufactured in a transgenic animal, such as a mammal or a bird, under control of a regulatory element such that expression of the ectopic EGF can be directed to a convenient production medium, such as production in the milk of a transgenic animal which is a mammal, or production in the albumin of an egg of a transgenic animal which is a bird. Further, the endogenous gene of the transgenic animal can be inactivated by a knock out mutation, such that the human modified EGF is the sole EGF produced in the transgenic animal. See, U.S. patent numbers 6,242,666, issued June 5, 2001, and 6,271,436, issued Aug. 7, 2001.

The term "ectopic" as used here and in the claims refers to a gene isolated from a cell of one type of organism, e.g., a human cell, and which following genetic engineering technologies has been introduced into a cell of a different type of organism, such as a yeast, or a heterologous mammal such as a pig.

"Optimum codon usage" as used here and in the claims refers to adjusting the nucleotide sequence of a recombinant engineered gene to reflect differences in frequencies of the plurality of different trinucleotide codons encoding the same amino acid that are found in the genes within genomes of different organisms. All amino acids other than trp and met can be encoded by a plurality of codons. More particularly, the term optimum codon usage refers to a procedure whereby an engineered nucleotide sequence encoding a protein or polypeptide, and intended for insertion into a heterologous cell for production of the protein by in vivo methods in that cell, is altered in such a way from the nature identical gene that substantially the same natural amino acid sequence is encoded using different codons. The different set of codons as designed in the engineered gene are chosen according to the frequency of usage in the production cell, so that elements of the translational apparatus of the production cell will not limit the yield of product.

As used herein, the terms "protein", "peptide", and "polypeptide" shall have the same meaning.

The contents of all cited patents and publication are hereby incorporated by reference herein. Although methods and materials similar or equivalent to those described herein can be used in the practice of the present invention, suitable methods and materials are described below. The invention having been fully described is illustrated by Examples and claims below, which are not intended to be further limiting.

### EXAMPLES

Example 1. Design and synthesis of an EGF gene having codons optimized for production, and amino acid sequence modifications for enhanced stability.

A foreign gene can be cloned and expressed in the industrially suitable yeast, *Pichia pastoris*, using a vector obtained, for example, from Invitrogen, Carlsbad, CA. A kit containing cells, vectors, and media can be purchased. Protocols are available from Invitrogen. Optimization of codons in *P. pastoris* can be determined by methods described in U.S. patent number 5,827,684, issued Oct. 27, 1998, and is available as a commercial service from Aptamer, Inc. (Rockville, MD).

The amino acid sequence at the C-terminus of full length EGF is, from residues 48 to 53, KWWELR, using the one letter code defined above (see Fig. 1 panel A, and SEQ ID No: 1). Surprisingly, a combination of a deletion of at least one amino acid, and replacement of at least one amino acid, of the 5 C-terminus amino acids, resulted in a modified form of EGF that retains its biological activity and is resistant to proteolysis. Modifications of these amino acids include: deletion or replacements of the basic amino acids at positions 48 (which in natural EGF is lysine, K) and 53 (arginine, R), and of the aromatic amino acids at position 49 (tryptophan, W) and 50 (tryptophan, W), and the aliphatic amino acid at position 52 (leucine, L).

These amino acids are here identified as target substrates for various exo and endo proteases of the type that can be found in gastric fluid, in circulation, and in the culture media of production microorganisms such as *P. pastoris*. The peptide bonds between amino acids of the carboxy terminus of nature identical EGF are subject to digestion by several proteases and peptidases of known specificity. For example, the bond between residues 52 and 53, leu-arg, is subject to carboxypeptidase B activity; the bonds between each of 49-50 (trp-trp) and 50-51 (trp-glu) are subject to chymotrypsin digestion; and the bond between 48-49 (lys-trp) is subject to trypsin digestion. (See "Proteolytic Enzymes", Ed. Beynon, R. and Bond J., IRL Press at Oxford, Appendix II, p. 232.) While additional proteases have other preferences or



can be non-specific, neutral amino acids such as gln are not often a preferred substrate. An exception is the protease subtilisin, a product of the Gram positive soil bacterium *Bacillus subtilis*. For this reason, production of modified EGF having neutral or acid residues as described herein, in the bacterium *B. subtilis* is not preferred.

An embodiment of the invention is a composition which is a peptide comprising the amino acid sequence of human EGF having a length of 51 amino acids, in which the residue at position 51 is an amino acid other than glutamic acid, for example, asparagine (identified as EGF51N). A related embodiment is a nucleotide sequence encoding a gene for the modified EGF composition EGF51N (SEQ ID NO: 2), as shown in Figure 1 panel B. This nucleotide sequence has been designed to be inserted into a vector such as an expression vector in an industrially acceptable production organism, for example, *Pichia pastoris*, for production of the composition, for use as a therapeutic agent. Similarly, EGF51A, EGF51Q, and EGF51S are further embodiments of the modified EGF forms. These modified forms are resistant to proteolysis as described herein (See Fig. 1 panels B, C, and D, respectively).

The design herein of a gene sequence encoding the modified EGF further uses codons that are optimal for recognition during protein synthesis in the production organism. These codons are substituted for the codons found in the natural human nucleotide sequence encoding EGF, in order to provide higher yields during growth of cells of the organism.

Further, data herein (*infra*) show that the design of the modified EGF in order to obtain resistance to proteolysis, for example, deleting two terminal amino acids at positions 52 and 53, and substituting asparagine for the naturally occurring glutamic acid at position 51, was successful in meeting this objective. Further, the modified form of EGF51N was found herein to have the desired biological activities, in particular, the ability to stimulate islet neogenesis to provide insulin in streptozotocin-induced diabetic rats, as shown in examples below.

Example 2. Yield of EGF from a recombinant organism using a synthetic gene having codons optimized for that organism

The gene sequence shown in Fig. 1 used to produce EGF51N in *Pichia pastoris*, gave surprisingly high yields of the product in the fermentation production medium (see Fig. 2). The yield was comparable to that previously reported (U.S. patent number 5,102,789 issued Apr. 7, 1992).

Example 3. Resistance of EGF51N to proteolysis.

Determination of the molecular weight of the product and analysis of peptides obtained by digestion of the EGF51N confirmed that a single species having a molecular

weight that conformed to the predicted molecular weight, and having a deletion of two amino acids compared to natural EGF, confirmed the identity of the material, and the production of a single species of EGF protein. During the time course of production of EGF51N, it was observed that the high yield of EGF obtained was characterized as a single molecular weight species (see Fig. 2). Thus, no significant proteolysis of the modified hEGF was observed in the fermentation medium.

Table 2 shows data obtained using another criterion indicating the resistance of EGF51N to protease activity, i.e., resistance *in vitro* to digestion with the enzyme carboxypeptidase B. These data show that between six and 30-fold more of EGF51N remains active following a course of incubation with carboxypeptidase B, compared to that of natural human EGF53. These data confirm that design of the C-terminus to eliminate residues that are substrates for enzyme digestion is effective in protecting the EGF during microbial production, and during exposure to other enzymes such as carboxypeptidase B in an environment such as in a subject.

Table 2. Resistance of EGF51N to carboxypeptidase B.

Relative enzyme concentration	% EGF53 remaining	% EGF51N remaining
0.01	75	75
0.1	1	30
1 (excess enzyme)	1	6

The relative enzyme concentration is the ratio of carboxypeptidase to EGF.

#### Example 4. Biological activity of EGF51N

EGF51N produced as above was used in an assay to measure to stimulation of islet neogenesis *in vivo*, as measured by a lower blood concentration treated animals to a subsequent challenge by glucose, using diabetic rats. Thus hEGF was found to have potency that is equivalent to the positive control TGF $\alpha$ , an alternative EGF receptor ligand (U.S. patent number 6,288,301, issued Sept. 11, 2001) as shown in Fig. 3.

Further, the modified EGF51N was found to be capable of inhibiting gastric acid secretion *in vivo* in anaesthetized gastric fistula rats. After intravenous bolus injection of 8 $\mu$ g EGF51N, the onset of inhibition of gastric acid production was found to have occurred within 10-20 minutes after injection, with a duration of inhibition of 20-50 minutes. This time of

onset and duration of inhibition of acid secretion are comparable to that of the positive control, commercially obtained full length TGF- $\alpha$ .

These data indicate that the C-terminus modifications embodied in the EGF forms described herein function both to protect EGF from proteolysis, and to maintain at least two different therapeutically important biological activities, inhibition of acid secretion and islet neogenesis therapy.

Example 5. Expression of EGF51N in mycelia of *Streptomyces lividans*

Protoplasts of *S. lividans* were prepared by standard methods (Hopwood et al., 1985, Genetic manipulation of Streptomyces" Norwich, England; The John Innes Foundation), and were transformed with a plasmid carrying the gene encoding EGF51N. The *Streptomyces* vector pCAN46 was obtained from CANGENE (Winnipeg, Manitoba). The plasmid uses a *Streptomyces* -E. coli shuttle vector carrying the pIJ101 replicon, which is a high copy number multicopy plasmid (see U.S. patent number 4,745,056) that replicates in a large variety of species of a host streptomycete cell. This plasmid carries a selectable antibiotic resistance gene encoding thiostrepton resistance. The expression of EGF51N uses the promoter originating from the *S. fradiae* gene encoding aminoglycoside phosphotransferase (aph) for protein expression. The construct further uses a genetic fusion of the amino terminal asparagine of EGF51N to the protease B signal sequence from *S. griseus* to obtain protein secretion.

Significant useful production of EGF51N was observed from mycelial growth of the strain carrying this construct.

What is claimed is:

1. A composition comprising an amino acid sequence of length X, X being an integer that is at least 48 and not more than 53, such sequence (i) being substantially homologous to a portion of SEQ ID NO: 1 from position 1 to position X-1 of SEQ ID NO: 1, and (ii) having at position X an amino acid residue different from that found in SEQ ID NO: 1.
2. A composition according to claim 1, wherein the amino acid residue at position X is a neutral amino acid.
3. A composition according to claim 1, wherein the amino acid residue at position X is asparagine.
4. A composition according to claim 3, wherein X is position 51.
5. A composition according to claim 1, having increased resistance to proteolysis in comparison with that of SEQ ID NO: 1.
6. A composition according to claim 5, wherein the biological activity is at least 75% of that of SEQ ID NO: 1.
7. A composition according to claim 6, wherein the biological activity is at least 90% of that of SEQ ID NO: 1.
8. A composition according to claim 6 or 7, wherein the biological activity is selected from the group consisting of: mitogenesis, cytoprotection, inhibition of acid secretion, growth of a tissue precursor cell, differentiation of a tissue precursor cell, and growth and differentiation of a tissue precursor cell.
9. A composition according to claim 8, wherein mitogenesis is determined by rate of mitosis of epithelial cells.
10. A composition according to claim 8, wherein acid secretion is determined in a gastric fistulated animal.
11. A composition according to claim 8, wherein differentiation is determined by islet neogenesis or mucosal cell formation.
12. A composition according to claim 1, wherein X is a hydrophobic amino acid.
13. A composition according to claim 1, wherein X is a charged amino acid.
14. A composition according to claim 13, wherein X is a negatively charged amino acid.
15. A composition according to claim 2, wherein X is selected from glutamine, alanine, and serine.

16. A composition comprising an amino acid sequence substantially homologous to SEQ ID NO: 1, wherein the amino acid residue at position 51 of SEQ ID NO: 1 is an amino acid other than glutamic acid.

17. A composition according to claim 16, wherein the biological activity of the composition is at least 50% of that of SEQ ID NO: 1.

18. A composition according to claim 17, wherein the biological activity of the composition is at least 75% of that of SEQ ID NO: 1.

19. A composition according to claim 18, wherein the biological activity of the composition is at least 90% of that of SEQ ID NO: 1.

20. A composition according to claim 16, the composition having a biological activity substantially equivalent to that of SEQ ID NO: 1.

21. A composition according to claim 16, wherein the amino acid residues at positions 1-50 are at least 75% identical to that of SEQ ID NO: 1.

22. A composition comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 2, 3 and 4.

23. A composition comprising an amino acid sequence as shown in SEQ ID NO: 2.

24. A polypeptide of 51 amino acids in length, wherein residues 1-50 are substantially homologous to the amino acid sequence as shown in SEQ ID NO: 1, and residue 51 is an asparagine residue.

25. A polypeptide according to claim 24, having increased resistance to proteolysis in comparison with that of SEQ ID NO: 1.

26. A polypeptide according to claim 24, wherein at least one of residues 1-50 is a conservative substitution of an amino acid in the sequence as shown in SEQ ID NO: 1.

27. A polypeptide according to claim 24, further comprising a deletion of at least one of residues selected from positions 1-5 as shown in SEQ ID NO: 1.

28. A polypeptide according to claim 26, having at least 50% of a biological activity of human EGF as shown in SEQ ID NO: 1.

29. A composition comprising a human epidermal growth factor (EGF) having an amino acid sequence substantially homologous to that of at least positions 1-47 as shown in SEQ ID NO: 1, and having at least one amino acid replacement at positions 48-53 of the EGF carboxy terminus, the amino acid sequence being more stable to proteolysis than that of SEQ ID NO: 1.

30. A composition according to claim 29, comprising the amino acid sequence for

residues at positions 1-50 substantially as shown in SEQ ID NO: 1, and in which the residue at position 51 is an amino acid other than glutamic acid.

31. A composition according to claim 30, in which the residue at position 51 is selected from the group consisting of asparagine, glutamine, alanine, and serine.

32. A composition according to claim 30, in which the residue at position 51 is asparagine.

33. A composition according to claim 30, wherein at least 75% of the amino acids at positions 1-50 are as shown in SEQ ID NO: 1.

34. A composition according to claim 30, wherein the biological activity of the composition is at least 50% of that shown in SEQ ID NO: 1.

35. A composition according to claim 34, wherein the biological activity is selected from the group consisting of mitogenesis, cytoprotection, inhibition of acid secretion, growth of a tissue precursor cell, differentiation of a precursor cell, and growth and differentiation of a precursor cell.

36. A composition according to claim 35, wherein mitogenesis is determined by rate of mitosis of epithelial cells.

37. A composition according to claim 35, wherein acid secretion is determined in a gastric fistulated animal.

38. A composition according to claim 35, wherein differentiation is determined by islet neogenesis or mucosal cell formation.

39. A pharmaceutical composition comprising an effective dose of the composition according to any of claims 1, 16 and 29 in a pharmaceutically acceptable carrier.

40. A pharmaceutical composition according to claim 39, further comprising an additional therapeutic agent.

41. A pharmaceutical composition according to claim 40, wherein the additional therapeutic agent is a growth factor receptor ligand.

42. A pharmaceutical composition according to claim 41, wherein the ligand is a growth factor.

43. A pharmaceutical composition according to claim 42, wherein the ligand is a gastrin/cholecystokinin receptor ligand.

44. A nucleotide sequence encoding a composition according to any of 1, 16, and 29, the nucleotide sequence having codons adjusted for optimum usage in an industrially acceptable production organism.

45. A nucleotide sequence according to claim 44, wherein the industrially acceptable production organism is a yeast.

46. A nucleotide sequence according to claim 45, wherein the yeast is *Pichia pastoris*.

47. A polynucleotide having a nucleotide sequence encoding a polypeptide of 51 residues in length and having a biological activity of human EGF, the sequence containing codons that are optimized for expression in a species of *Pichia*, and having an amino acid at the carboxyl terminus capable of conferring resistance to proteolysis.

48. A recombinant strain of *Pichia* carrying a nucleotide sequence as shown in SEQ ID NO: 5.

49. A recombinant strain of *Pichia* capable of producing the amino acid sequence as shown in SEQ ID NO: 2.

50. A nucleotide sequence encoding an amino acid sequence as shown in SEQ ID NO: 2.

51. A method of obtaining a composition comprising an amino acid sequence of length X, X being an integer that is at least 48 and not more than 53, such sequence (i) being substantially homologous to the portion of SEQ ID NO: 1 from position 1 to position X-1 of SEQ ID NO: 1, and (ii) having at position X an amino acid residue different from that found in SEQ ID NO: 1, the method comprising:

designing a gene encoding the composition, having codons selected for optimum usage in an industrially acceptable organism; and

producing the composition in the organism by fermentation of the organism.

52. A use of a composition for manufacture of a treatment for a subject in need of regeneration of a tissue, the therapeutic composition comprising an amino acid sequence of length X, X being an integer that is at least 48 and not more than 53, such sequence (i) being substantially homologous to the portion of SEQ ID NO: 1 from position 1 to position X-1 of SEQ ID NO: 1, and (ii) having at position X an amino acid residue different from that found in SEQ ID NO: 1, the method comprising:

obtaining the composition by the method according to claim 51; and

administering to the subject the composition in an amount sufficient to effect regeneration of precursor cells, as a treatment of the subject for regeneration of the tissue.

53. A use according to claim 52, wherein administering the composition is further administering an additional therapeutic agent.

54. A use according to claim 53, wherein the additional therapeutic agent is a growth factor receptor ligand.

55. A use according to claim 54, wherein the ligand is a growth factor.

56. A use according to claim 54, wherein the ligand is a gastrin/cholecystokinin receptor ligand.

57. A use according to claim 52, wherein X is the integer 51.

58. A use according to claim 57, wherein the amino acid residue at X is a neutral amino acid.

59. A use according to claim 58, wherein the amino acid residue is asparagine.

60. A use according to claim 52, wherein the composition is more resistant to proteolysis than the composition having a sequence as shown in SEQ ID NO: 1.

61. A use according to claim 52, wherein the subject is in need of islet cell regeneration.

62. A use according to claim 52, wherein the subject is in need of mucosal cell regeneration.

63. A use according to claim 52, wherein the subject has diabetes.

64. A method of obtaining a modified amino acid sequence of an EGF that is more resistant to proteolysis than a nature identical EGF and that substantially retains a biological activity, the method comprising:

identifying at least one residue of a sequence of the nature identical EGF that is subject to proteolysis by an industrially useful organism;

designing the modified amino acid sequence in which the residue identified as subject to proteolysis is deleted or substituted with a different amino acid; and

providing the modified amino acid sequence and testing proteolysis in comparison to the nature identical EGF, to obtain a modified EGF that is more resistant to proteolysis than the nature identical EGF.

65. A method according to claim 64, in which the nature identical EGF is from a human.



66. A method according to claim 65, in which designing the modified amino acid sequence is: substituting at least one amino acid for at least one carboxy terminal residue, to obtain a new carboxy terminal amino acid sequence, such that the modified EGF is more resistant to proteolysis.

67. A method according to claim 65, wherein identifying the at least one residue that is subject to proteolysis is determining a site of proteolysis during production of the nature identical EGF in a culture of a recombinant cell of the industrially useful organism.

68. A method according to claim 65, wherein following designing the modified amino acid sequence and prior to providing the modified EGF, further includes designing a nucleotide sequence encoding the modified amino acid sequence, the nucleotide sequence having codons selected for optimal usage in the industrially useful organism.

69. A method according to claim 68, wherein following designing the modified nucleotide sequence and prior to providing the modified EGF, further includes incorporating the modified nucleotide sequence into a vector, and transforming the vector into a cell of the industrially useful organism.

70. A method according to claim 69, wherein providing the modified EGF is providing a protein having an amino acid sequence as shown in SEQ ID NO:2.

71. A method according to claim 70, wherein the industrially useful organism is *Pichia pastoris*.

72. A kit comprising at least one unit dosage of the composition of any of claims 1, 16, and 29, in a pharmaceutically acceptable carrier.

73. A kit according to claim 72, further comprising an additional therapeutic agent.

74. A kit according to claim 73, wherein the additional therapeutic agent is a growth receptor ligand.

75. A kit according to claim 69, wherein the growth receptor ligand is a gastrin/cholecystokinin receptor ligand.

76. A kit according to claim 75, wherein the ligand is gastrin.

77. A kit according to claim 72, wherein the unit dosage is sufficient for treatment of a subject in need of regeneration of a tissue.

78. A kit according to claim 77, wherein the tissue is a pancreatic islet or a gastric mucosa.

79. A recombinant strain of *Pichia* obtained according to the method of claim 71.

80 A recombinant strain of *Streptomyces* carrying a nucleotide sequence as shown in SEQ ID NO: 5.

81. A recombinant strain of *Streptomyces* capable of producing the amino acid sequence as shown in SEQ ID NO: 2.

82. A recombinant strain of *Streptomyces* obtained according to the method of claim 71.

1/6

AAC TCT GAC TCC GAA TGT CCA TTG TCT CAC  
asn ser asp ser glu cys pro leu ser his

GAC GGT TAC TGT TTG CAC GAC GGT GTT TGT  
asp gly tyr cys leu his asp gly val cys

ATG TAC ATC GAA GCT TTG GAC AAG TAC GCT  
met tyr ile glu ala leu asp lys tyr ala

TGT AAC TGT GTT GTC GGT TAC ATC GGT GAA  
cys asn cys val val gly tyr ile gly glu

AGA TGT CAA TAC AGA GAC TTG AAG TGG TGG  
arg cys gln tyr arg asp leu lys trp trp

(G) AA T TG A GA T AA  
deletion asn stop

PANEL A  
**FIG. 1A**

2/6

AAC TCT GAC TCC GAA TGT CCA TTG TCT CAC  
asn ser asp ser glu cys pro leu ser his

GAC GGT TAC TGT TTG CAC GAC GGT GTT TGT  
asp gly tyr cys leu his asp gly val cys

ATG TAC ATC GAA GCT TTG GAC AAG TAC GCT  
met tyr ile glu ala leu asp lys tyr ala

TGT AAC TGT GTT GTC GGT TAC ATC GGT GAA  
cys asn cys val val gly tyr ile gly glu

AGA TGT CAA TAC AGA GAC TTG AAG TGG TGG  
arg cys gln tyr arg asp leu lys trp trp

GCT TGA GAT AA  
ala stop

PANEL B  
**FIG. 1B**

3/6

AAC TCT GAC TCC GAA TGT CCA TTG TCT CAC  
asn ser asp ser glu cys pro leu ser his

GAC GGT TAC TGT TTG CAC GAC GGT GTT TGT  
asp gly tyr cys leu his asp gly val cys

ATG TAC ATC GAA GCT TTG GAC AAG TAC GCT  
met tyr ile glu ala leu asp lys tyr ala

TGT AAC TGT GTT GTC GGT TAC ATC GGT GAA  
cys asn cys val val gly tyr ile gly glu

AGA TGT CAA TAC AGA GAC TTG AAG TGG TGG  
arg cys gln tyr arg asp leu lys trp trp

CAA TGA GAT AA  
gln stop

PANEL C  
FIG. 1C

4/6

AAC TCT GAC TCC GAA TGT CCA TTG TCT CAC  
asn ser asp ser glu cys pro leu ser his

GAC GGT TAC TGT TTG CAC GAC GGT GTT TGT  
asp gly tyr cys leu his asp gly val cys

ATG TAC ATC GAA GCT TTG GAC AAG TAC GCT  
met tyr ile glu ala leu asp lys tyr ala

TGT AAC TGT GTT GTC GGT TAC ATC GGT GAA  
cys asn cys val val gly tyr ile gly glu

AGA TGT CAA TAC AGA GAC TTG AAG TGG TGG  
arg cys gln tyr arg asp leu lys trp trp

TCT TGA GAT AA  
ser stop

PANEL D  
FIG. 1D

5/6

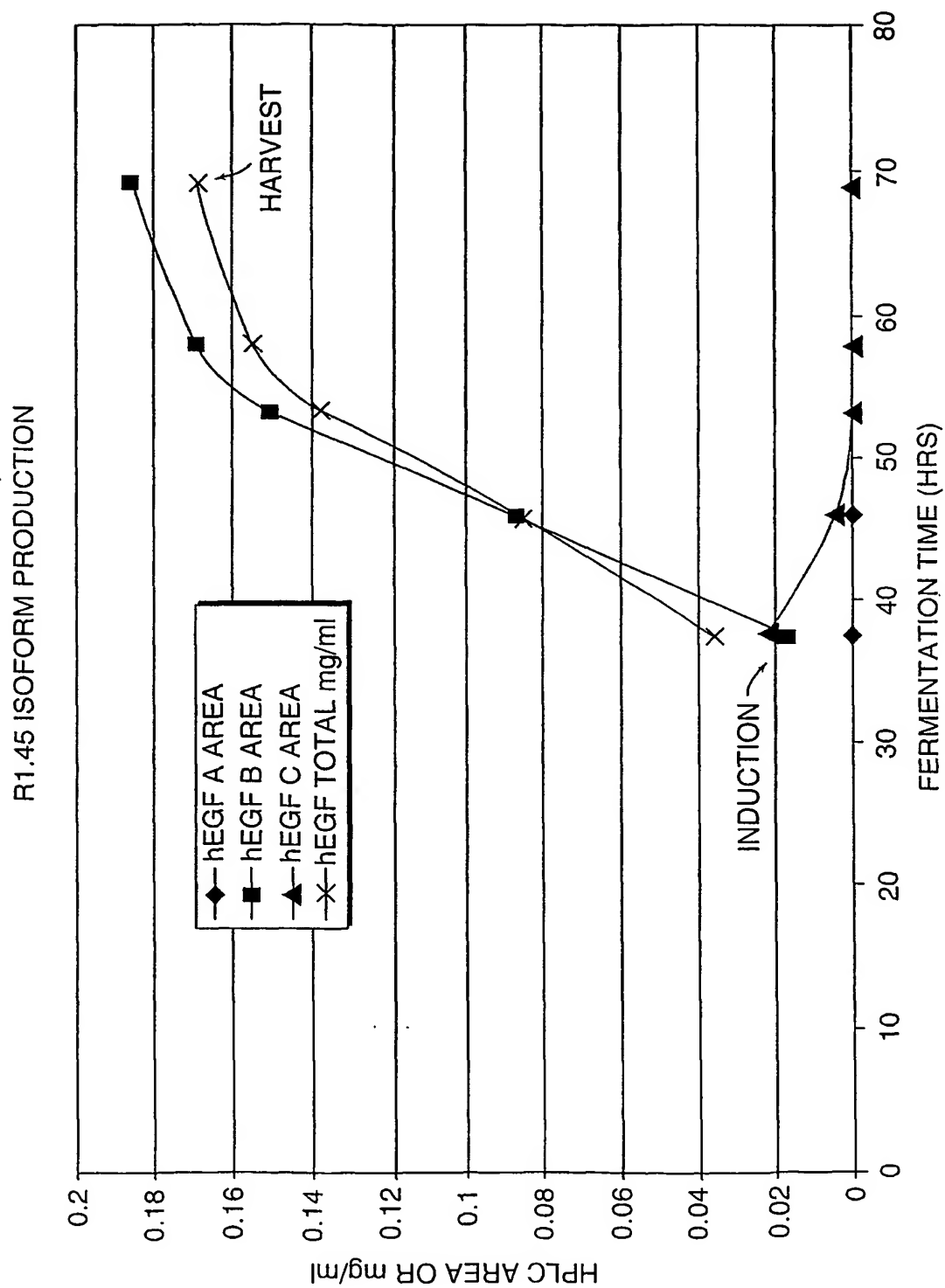


FIG. 2

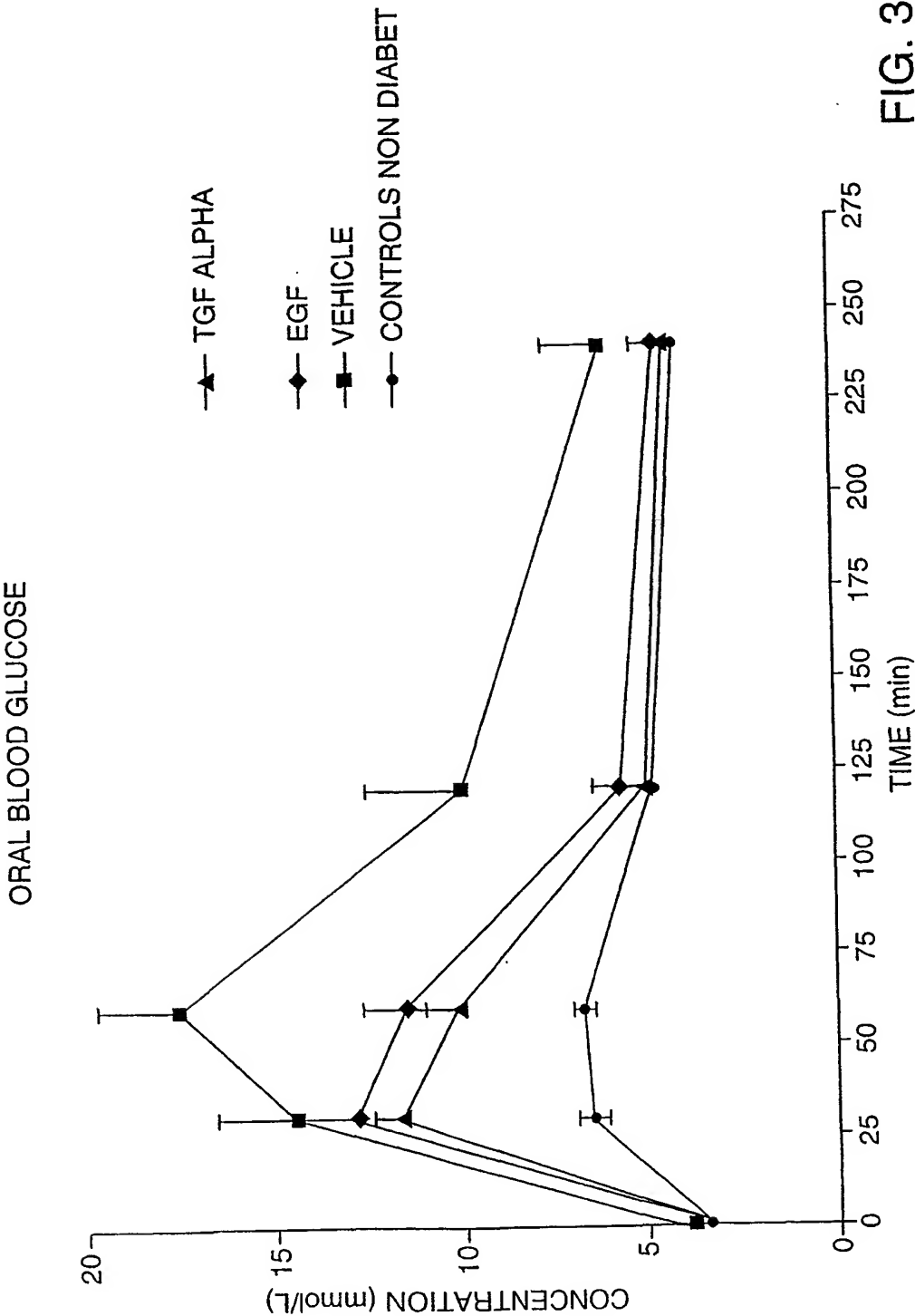


FIG. 3



## SEQUENCE LISTING

<110> Magil, Sheila G.  
 Jones, Susan D.  
 Pace, Gary W.  
 Brand, Stephen J.

<120> Novel Epidermal Growth Factor Protein and  
 Gene, and Methods of Use Therefor

<130> 24492-005-061

<140> Not Yet Assigned

<141> 2002-10-22

<150> US 10/000,840

<151> 2001-10-23

<160> 9

<170> PatentIn Ver. 2.1

<210> 1

<211> 53

<212> PRT

<213> Human EGF (epidermal growth factor)

<400> 1

```

Asn Ser Asp Ser Glu Cys Pro Leu Ser His Asp Gly Tyr Cys Leu His
 1      5      10
Asp Gly Val Cys Met Tyr Ile Glu Ala Leu Asp Lys Tyr Ala Cys Asn
      20      25      30
Cys Val Val Gly Tyr Ile Gly Glu Arg Cys Gln Tyr Arg Asp Leu Lys
      35      40      45
Trp Trp Glu Leu Arg
      50
  
```

<210> 2

<211> 51

<212> PRT

<213> Recombinant human EGF51N

<400> 2

```

Asn Ser Asp Ser Glu Cys Pro Leu Ser His Asp Gly Tyr Cys Leu His
 1      5      10
Asp Gly Val Cys Met Tyr Ile Glu Ala Leu Asp Lys Tyr Ala Cys Asn
      20      25      30
Cys Val Val Gly Tyr Ile Gly Glu Arg Cys Gln Tyr Arg Asp Leu Lys
      35      40      45
Trp Trp Asn
      50
  
```

<210> 3

<211> 51

<212> PRT

<213> Recombinant human EGF51A

<400> 3

```

Asn Ser Asp Ser Glu Cys Pro Leu Ser His Asp Gly Tyr Cys Leu His
 1      5      10
Asp Gly Val Cys Met Tyr Ile Glu Ala Leu Asp Lys Tyr Ala Cys Asn
      20      25      30
Cys Val Val Gly Tyr Ile Gly Glu Arg Cys Gln Tyr Arg Asp Leu Lys
      35      40      45
Trp Trp Ala
      50
  
```

<210> 4  
 <211> 51  
 <212> PRT  
 <213> Recombinant human EGF51Q

<400> 4  
 Asn Ser Asp Ser Glu Cys Pro Leu Ser His Asp Gly Tyr Cys Leu His  
 1 5 10 15  
 Asp Gly Val Cys Met Tyr Ile Glu Ala Leu Asp Lys Tyr Ala Cys Asn  
 20 25 30  
 Cys Val Val Gly Tyr Ile Gly Glu Arg Cys Gln Tyr Arg Asp Leu Lys  
 35 40 45  
 Trp Trp Gln  
 50

<210> 5  
 <211> 51  
 <212> PRT  
 <213> Recombinant human EGF51S

<400> 5  
 Asn Ser Asp Ser Glu Cys Pro Leu Ser His Asp Gly Tyr Cys Leu His  
 1 5 10 15  
 Asp Gly Val Cys Met Tyr Ile Glu Ala Leu Asp Lys Tyr Ala Cys Asn  
 20 25 30  
 Cys Val Val Gly Tyr Ile Gly Glu Arg Cys Gln Tyr Arg Asp Leu Lys  
 35 40 45  
 Trp Trp Ser  
 50

<210> 6  
 <211> 161  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> DNA encoding EGF51N

<400> 6  
 aactctgact ccgaatgtcc attgtctcac gacggttact gtttgcacga cgggtgtttgt 60  
 atgtacatcg aagcttttga caagtacgct tgtaactgtg ttgtcggta catcggtgaa 120  
 agatgtcaat acagagactt gaagtgggtg aattgagata a 161

<210> 7  
 <211> 161  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> DNA encoding EGF51A

<400> 7  
 aactctgact ccgaatgtcc attgtctcac gacggttact gtttgcacga cgggtgtttgt 60  
 atgtacatcg aagcttttga caagtacgct tgtaactgtg ttgtcggta catcggtgaa 120  
 agatgtcaat acagagactt gaagtgggtg gcttgagata a 161

<210> 8  
 <211> 161  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> DNA encoding EGF51Q

<400> 8  
aactctgact ccgaatgtcc attgtctcac gacggttact gtttgcacga cgggtgtttgt 60  
atgtacatcg aagcttttga caagtacgct tgtaactgtg ttgtcgggta catcggtgaa 120  
agatgtcaat acagagactt gaagtgggtg caatgagata a 161

<210> 9

<211> 161

<212> DNA

<213> Artificial Sequence

<220>

<223> DNA encoding EGF51S

<400> 9  
aactctgact ccgaatgtcc attgtctcac gacggttact gtttgcacga cgggtgtttgt 60  
atgtacatcg aagcttttga caagtacgct tgtaactgtg ttgtcgggta catcggtgaa 120  
agatgtcaat acagagactt gaagtgggtg tcttgagata a 161

THIS PAGE BLANK COPY